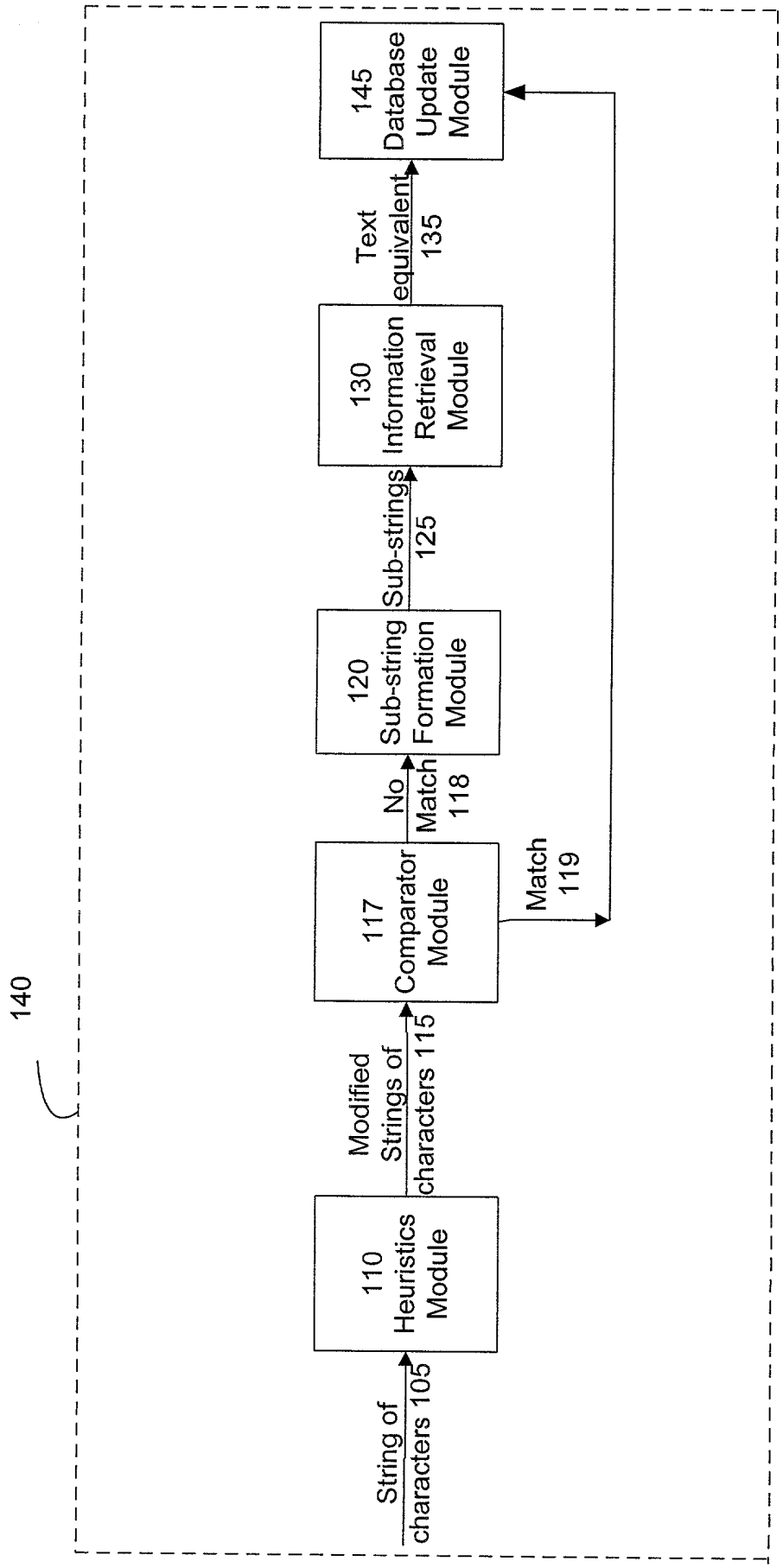
140
String of characters 105
110 Heuristics Module
Modified Strings of characters 115
117 Comparator Module
No Match 118
Match 119
120 Sub-string Formation Module
Sub-strings 125
130 Information Retrieval Module
Text equivalent 135
145 Database Update Module

FIG. 1

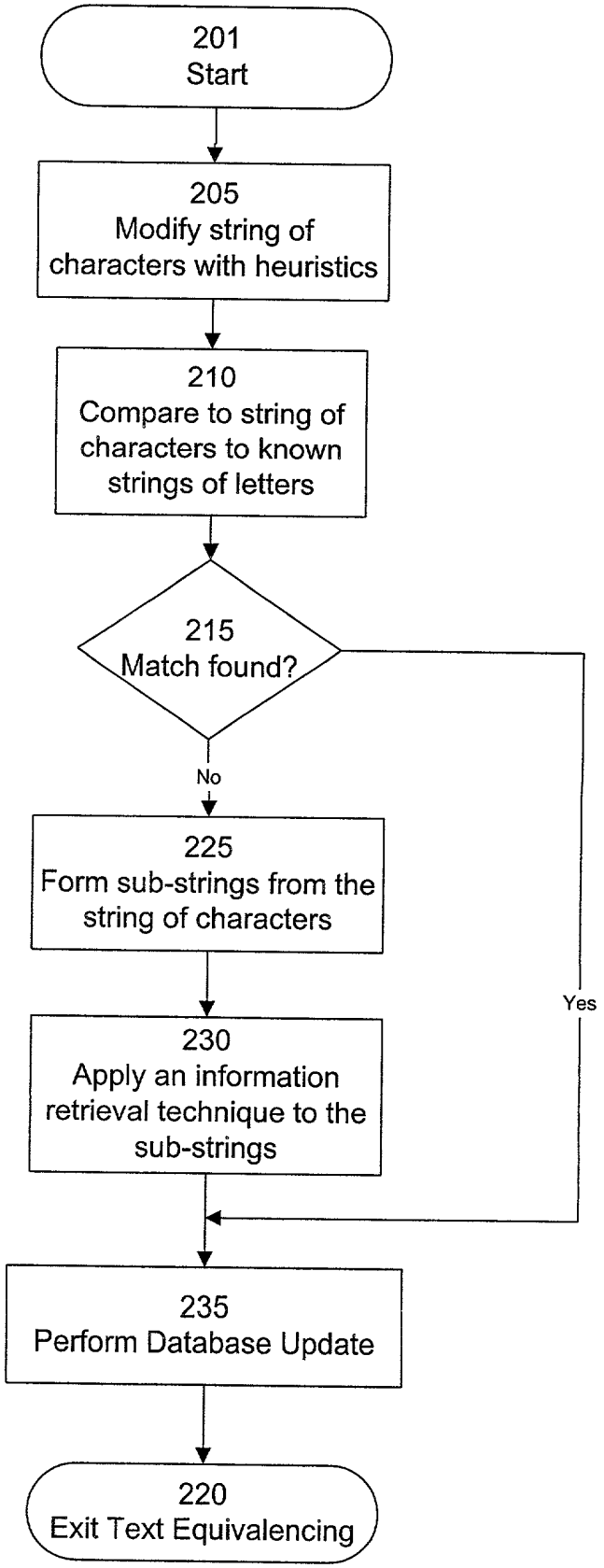
201
Start
205
Modify string of characters with heuristics
210
Compare to string of characters to known strings of letters
215
Match found?
No
Yes
225
Form sub-strings from the string of characters
230
Apply an information retrieval technique to the sub-strings
235
Perform Database Update
220
Exit Text Equivalencing

FIG. 2

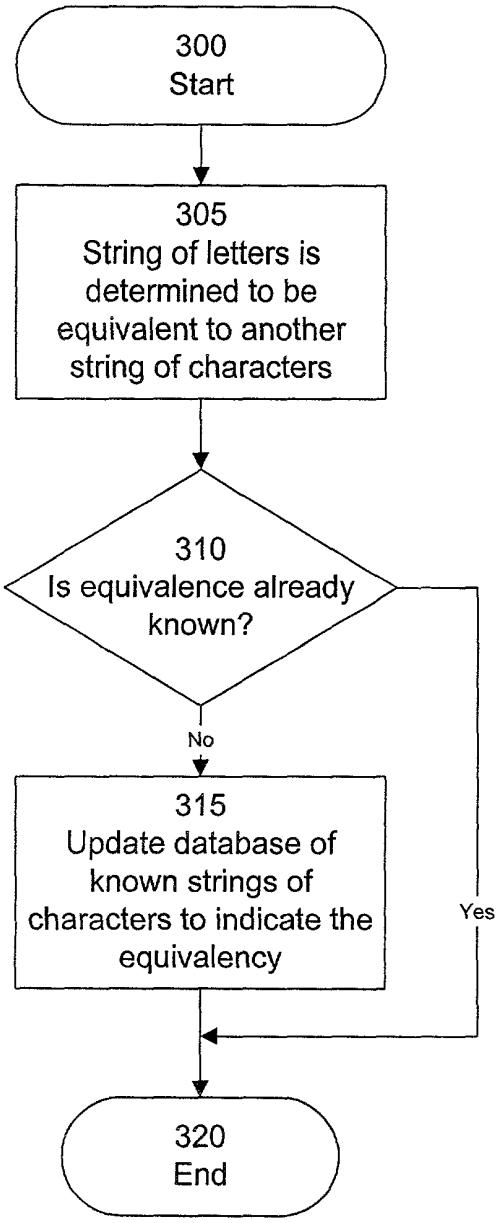
300
Start
305
String of letters is determined to be equivalent to another string of characters
310
Is equivalence already known?
No
Yes
315
Update database of known strings of characters to indicate the equivalency
320
End

FIG. 3

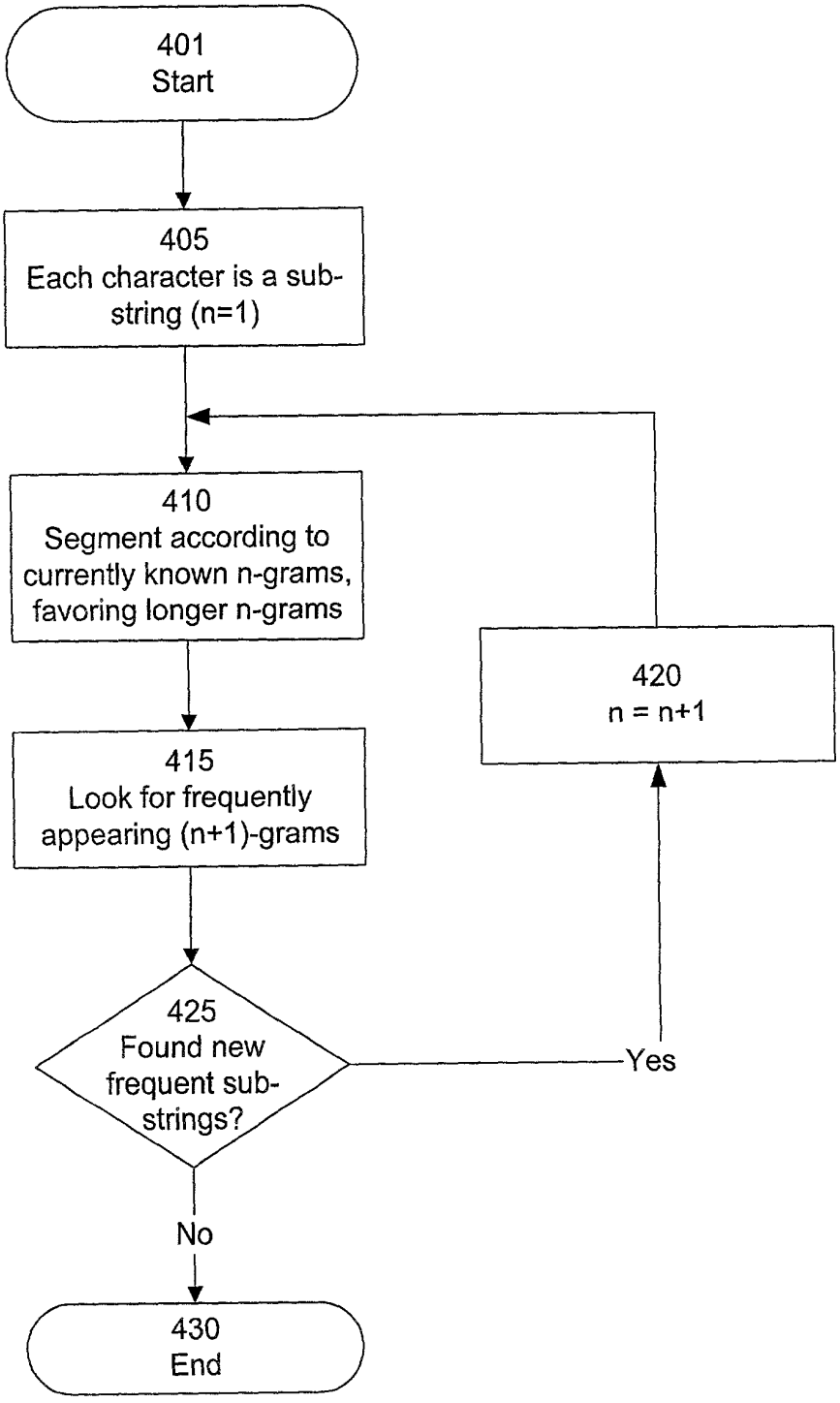
401
Start
405
Each character is a sub-string (n=1)
410
Segment according to currently known n-grams, favoring longer n-grams
415
Look for frequently appearing (n+1)-grams
425
Found new frequent sub-strings?
Yes
No
420
n = n+1
430
End

FIG. 4

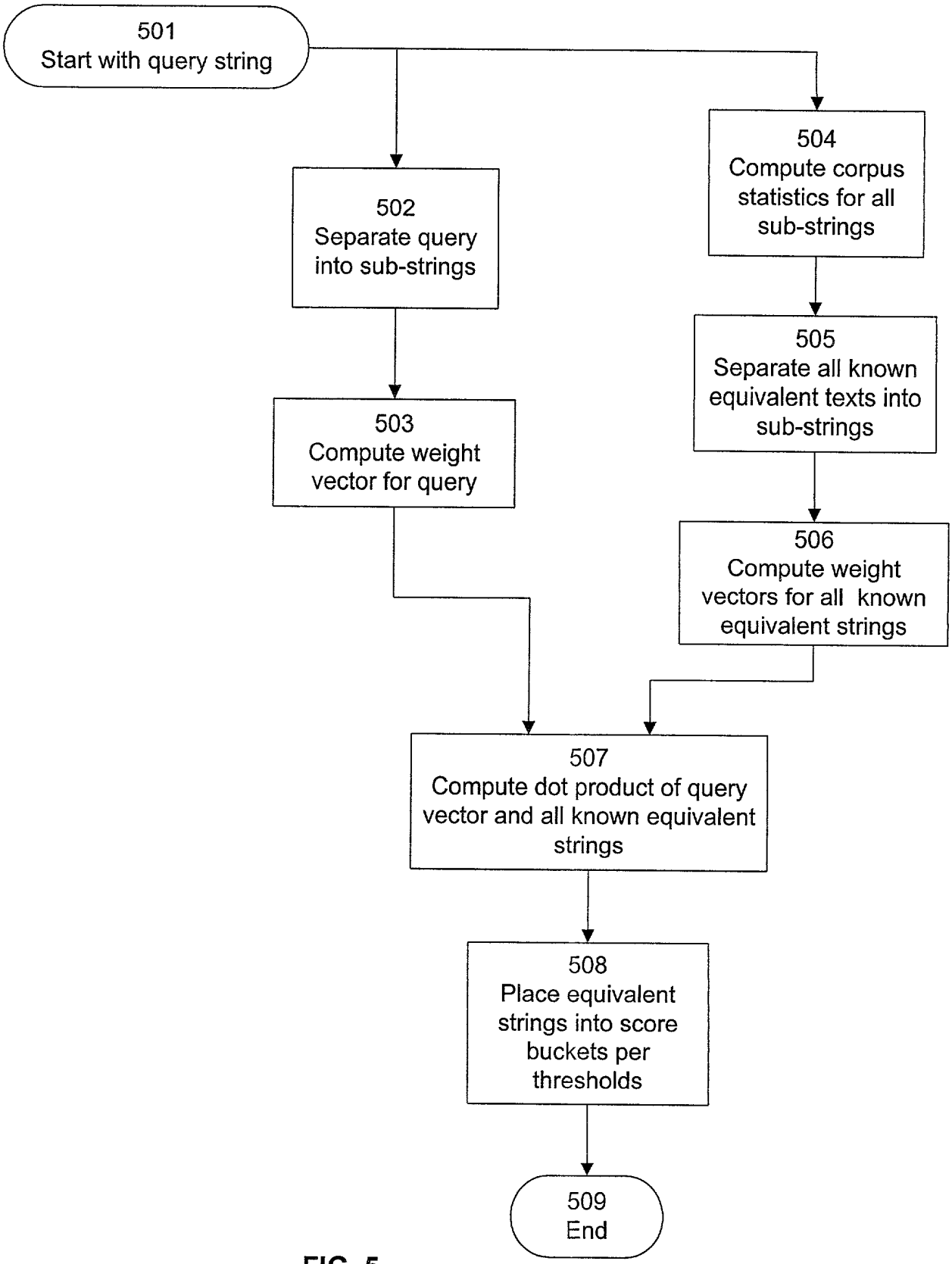
501
Start with query string
502
Separate query into sub-strings
503
Compute weight vector for query
504
Compute corpus statistics for all sub-strings
505
Separate all known equivalent texts into sub-strings
506
Compute weight vectors for all known equivalent strings
507
Compute dot product of query vector and all known equivalent strings
508
Place equivalent strings into score buckets per thresholds
509
End

FIG. 5